Quality and User Experience (2018) 3:1
https://doi.org/10.1007/s41233-017-0015-4

RESEARCH ARTICLE

The role of human influence factors on overall listening experience

Tim Walton[1,2] · **Michael Evans**[2]

Received: 8 August 2017

Abstract
Overall listening experience (OLE) is an evaluation measure specific to the evaluation of audio, which aims to include all possible factors that may influence listeners' ratings of stimuli. As with quality of experience in general, OLE ratings are user dependent. Previous research has shown that listeners can be categorised by how much their OLE is influenced by content and technical audio quality respectively. In this article, we expand on this knowledge by investigating correlations between a range of human influence factors and the degree to which a listener is influenced by content and technical audio quality. This was done by means of a web-based experiment involving 58 participants from a range of backgrounds. Results show that listener type is significantly correlated with a range of psychographic variables and that the attitudinal measure 'competence' is the most suitable variable to be used as a predictor of listener type. As well as these results having direct applications such as tailoring systems and services to the needs of specific user groups, the results presented add to the understanding of how human factors can influence quality of experience in general.

Keywords Overall listening experience · Quality of experience · Human influence factors · Psychographic · Attitudes · Demographics

Introduction

Subjective evaluation is a fundamental process in the advancement of multimedia services and systems, not least in the context of audio technology. In order to evaluate such technology in an ecologically valid manner, it is necessary to consider factors such as the type of content to be consumed with the technology, where the technology will be used and who will be using the technology. These considerations are central to the notion of quality of experience (QoE) and it can therefore be beneficial to take a quality of experience approach when evaluating audio technology.

One evaluation measure that has been introduced with the aim of evaluating audio technology with a QoE mindset is overall listening experience (OLE) [1]. This is an affective measure that is intended to include all possible factors that may influence listeners' ratings of stimuli, for example the song, lyrics, technical audio quality, the listeners mood and the reproduction system. As with QoE, by it's nature OLE is therefore user (or listener) dependent. This was highlighted in a study which showed that the relative influence of content and technical quality on OLE is very listener dependent; on the one hand some users are heavily influenced by content when making OLE judgements and, on the other hand, some users are heavily influenced by technical audio quality when making OLE judgements, with a continuum of users between [2].

In order to tailor systems and services to the appropriate audience, it would be beneficial to know what types of listeners are using the services and systems in question. For example, if it was known that the overall listening experience of a certain user group is highly influenced by technical audio quality, it would be desirable to provide them with the best quality available. Likewise, if it was known that the overall listening experience of a different user group is highly influenced by the content, it would be less problematic if the quality was reduced.

✉ Tim Walton
t.walton3@ncl.ac.uk

Michael Evans
michael.evans@bbc.co.uk

[1] Open Lab, Newcastle University, Newcastle upon Tyne, UK

[2] BBC Research and Development, Salford, UK

Published online: 13 January 2018

Fig. 1 Factors influencing QoE can be grouped into system, context and human influence factors. As represented in the figure, these groups of influence factors often overlap and together have a mutual impact on QoE. Adapted from [5]

In this paper, an experimental study is presented with the aim of identifying psychographic[1] variables that significantly influence whether a listener is heavily influenced by content or quality when making OLE ratings. As well as having direct applications, such as those mentioned above, this study provides insight into how human factors can influence quality of experience in general.

Related work

Human influence factors

Quality of experience can be subject to a range of factors that influence the human experience. Such 'influence factors' (IFs) can be grouped into *system*, *context* and *human* influence factors and, due to the complex and interrelated nature of QoE IFs, these groups of influence factors often overlap and together have a mutual impact on QoE [4]. This mutual impact of IFs on QoE is portrayed in Fig. 1. Of these three categories of IFs, the relationship between human influence factors (HIFs) and QoE is perhaps the least understood, due to the inherent complexity of HIFs and the lack of empirical evidence [5]. Human influence factors are defined as *"any variant or invariant property or characteristic of a human user"* where such characteristics *"describe the demographic and socio-economic back-ground, the physical and mental constitution, or the user's emotional state"* [4].

HIFs can be categorised into those that deal with low-level processing and those that deal with higher-level processing. Low-level HIFs relate to the physical, emotional and mental constitution of the user whereas higher-level HIFs relate to the understanding of stimuli and the associated interpretative and evaluative processes [6]. Examples of low-level HIFs include sensorial acuity, gender, age, emotions, mood and attention level, whereas examples of high-level HIFs include knowledge, skills, previous experiences, socio-cultural background, values and motivation.

Previous studies in the field of QoE have investigated a range of these HIFs. For example, Jumisko-Pyykkö and Hakkinen [7] investigated the impact of psychographic variables on the consumer-oriented quality assessment of mobile television. The studied variables were age, gender, education, professionalism, television consumption, experiences of different digital video qualities, and attitude towards technology. The results showed that quality evaluations were affected by almost all background factors. In a study by Wechsung et al. [8], it was shown that attitudes and mood are related to quality perceptions, however no link was found between personality traits and perceived quality. Other studies include those looking at the influence of mood and emotions [9–11], motivation [12] and expectations [13–15] on QoE.

With the exception of previous experiences and prior knowledge, the study of human influence factors in the field of audio evaluation is much more limited. Previous experiences and prior knowledge are typically used to distinguish between expert and 'naïve' listeners, for example in [16]. Expert listeners are used as reliable 'quality meters' who can identify small differences between stimuli, whereas naïve listeners provide results with more external validity. Quintero and Raake [17] however, investigated how factors beyond the level of prior knowledge of users affects perception of quality in the context of speech quality evaluation. Users were classified into six groups according to their demographic characteristics, their attitude towards adopting new technologies and socio-economic information. Significantly different quality ratings between these groups were found. Other studies include those looking at the influence of cultural backgrounds on timbre preferences [18], the influence of listeners' experience, age, and culture on headphone sound quality preferences [19] and various studies looking at the impact of language on quality perception of speech [20, 21].

In the study presented here we build on this previous work by investigating the influence of a range of human influence factors, both low- and high-level, on the impact of technical audio quality on the overall listening experience.

Overall listening experience

Overall listening experience is an affective measure recently used and defined by Schoeffler and Herre [1] that is inspired

[1] *Pyschographics* can be defined as "The study and classification of people according to their attitudes, aspirations, and other psychological criteria..." [3].

by the notion of QoE. The term is used to describe the degree of enjoyment whilst listening to audio, i.e. it is QoE defined specifically for the case of audio consumption. OLE and QoE are comparable in the sense that they both intend to take into account all possible factors that may influence a user's enjoyment. For both OLE and QoE these factors could include system influence factors such as the reproduction system, human influence factors such as mood, and context influence factors such as the listening environment. Whereas QoE is a concept that is applicable to a range of research areas, OLE is a specific application of QoE for the field of audio evaluation.

To assess OLE, participants are asked to rate stimuli on a five-star Likert scale taking everything into consideration that is important to them (e.g. quality, content etc.). Ratings are first given for reference conditions (i.e. unprocessed stimuli) and these act as a measure of how much participants like each song without taking any processing into account. These 'basic item ratings' are given through a multi-stimulus procedure so as to reduce floor and ceiling effects [22]. Secondly, the conditions to be tested (e.g. different reproduction methods) are rated and these are known as 'item ratings'. These are given through a single-stimulus procedure as such an approach is more representative of real-world listening scenarios [22]. Although the method combines both multi- and single-stimulus aspects, it has been shown that OLE ratings retrieved from a multi-stimulus procedure are consistent with those retrieved from a single-stimulus procedure and are thus comparable. With the basic item ratings and item ratings it is then possible to evaluate how much the different conditions influence the overall listening experience as well as evaluating to what extent listeners' ratings are influenced by the content and processing conditions respectively.

OLE has been used in a range of previous studies including investigations on the influence of timbral audio quality on OLE [23], the influence of up-/down-mixes on OLE [24], the influence of single-/multi-channel systems on OLE [25] and for the evaluation of 3D audio systems [26]. Furthermore, comparisons between OLE and basic audio quality have been made [27] in which it was seen that OLE can produce comparable results to basic audio quality.

As mentioned in the introduction, previous work has also highlighted the user dependent nature of OLE [2]. It was shown that for some listeners OLE is highly dependent upon technical audio quality, whereas for others, OLE is highly dependent upon content, with a continuum of listeners between. To study listener type, Schoeffler and Herre [2] suggest the approach of using correlation analysis, specifically Kendall rank correlation coefficients (Kendall's τ), in relation to given OLE ratings. In short, to determine how much the content influences each participant's OLE ratings, the correlation between their item ratings and basic item ratings is calculated. To determine how much the technical audio quality influences each participant's OLE ratings, the correlation between their item ratings and the various quality levels under study is calculated. Each listener is therefore described by a pair of correlation coefficients describing to what extent they are influenced by content and technical audio quality respectively. To date, factors that relate to, and can therefore be used to predict, listener type have not been studied or discussed in the literature. In this study, we build on this previous work by using the above approach in conjunction with a questionnaire in order to identify psychographic variables that significantly influence listener type.

Experimental procedure

This experiment was conducted as a web-based study. This was seen as appropriate as a large number of participants were required from a range of backgrounds and, furthermore, the differences between stimuli were not so small as to necessitate strict laboratory reproduction conditions. Moreover, a web-based approach leads to a higher external validity, which is an important consideration when evaluating quality of experience; participants listened to the content in a situation typical of their normal listening environment and with the technology that they would typically use.

The study was split into three sections—an online questionnaire to collect psychographic data and two online listening sessions. These are described in more detail in the following sections. It should be noted that a secondary aim of this experiment was to investigate the influence of binaural audio on OLE. Results concerning this objective have previously been published and, as such, segments of this chapter and the results chapter can also be found in [28].

Psychographic data collection

The psychographic data were collected by means of an online questionnaire, the overall form of which was inspired by [7]. The data collected can be roughly categorised into groups relating to demographics, experience and attitudes towards audio technology. Additionally, name and email were collected during each session for identification purposes.

Demographics

Data collected relating to demographics includes gender, age group, level of education (British system) and self-reported hearing normality.

Experience

To assess experience in the field of audio technology and specific experience relating to headphone usage and binaural audio experience, the following four questions were used.

- Select the statement that best describes the role of audio technology in your work and hobbies:
  - I study or work mainly in the field of audio technology
  - My work or hobbies involve some knowledge of audio technology
  - My work or hobbies are not related to audio technology
- Select the statement that best describes your headphone listening habits:
  - I listen to audio over headphones most days
  - I often listen to audio over headphones
  - I rarely listen to audio over headphones
  - I never listen to audio over headphones
- Select the statement that best describes your experience with binaural audio:
  - I have no experience of listening to binaural audio
  - I have limited experience of listening to binaural audio
  - I am experienced in listening to binaural audio
  - I'm not sure
- How many listening experiments have you previously participated in?
  - None
  - 1–5
  - 6–10
  - More than 10

Attitudes towards audio technology

A combination of two previously reported questionnaires was used to measure attitudes towards audio technology. The first of these, The Domain Specific Innovativeness (DSI) scale [29], has previously been used in a range of fields to measure consumer innovativeness, including studies related to quality assessment of mobile television [7]. In addition to this scale, parts of a questionnaire designed to measure technical affinity, known as the TA-EG [30], were used to measure competence and enthusiasm. This questionnaire was originally designed for use with German speakers and was therefore translated for this study. The complete list of statements to measure attitudes towards audio technology is as follows:

- Competence
  - I know most functions on the audio devices I own
  - *I struggle/would struggle to understand audio technology magazines*
  - I find it easy learning how to operate audio devices
  - I'm well versed in the field of audio technology
- Enthusiasm
  - I stay informed about audio technology, even if I don't intend to make a purchase
  - I love owning new audio technology
  - I get excited when a new device related to audio technology is brought to market
  - I like to go into specialist retailers for audio technology
  - I enjoy trying out audio technology
- Domain specific innovativness
  - *In general, I am among the last in my circle of friends to buy new audio technology when it appears*
  - If I heard that a new item of audio technology was available to purchase, I would be interested enough to buy it
  - *Compared to my friends I don't own much audio technology*
  - *In general, I am the last in my circle of friends to know about the latest audio technology*
  - I will not buy new audio technology if I haven't tried it yet
  - I like to buy new audio technology before other people do

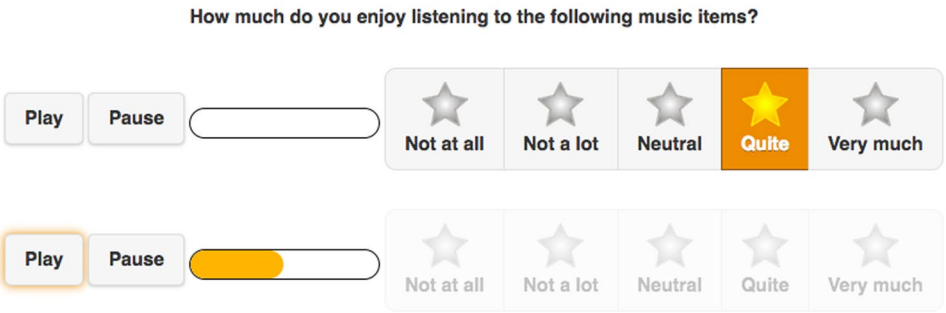


Fig. 2 User interface for OLE ratings

These claims were presented in a continuous list without the headings shown above. Participants were instructed to 'rate your attitude towards audio technology with the following statements' with ratings being made on a five-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. The statements in italic type are negatively phrased and the corresponding scores must therefore be reversed for analysis.

Procedure of listening sessions

As mentioned previously, the evaluation measure used in this study is OLE. To assess OLE, participants are asked to rate stimuli on a five-star Likert scale taking everything into consideration that is important to them. Ratings are first given for reference conditions and these act as a measure of how much participants like each song without taking any processing into account. These ratings are known as 'basic item ratings' (BIRs). Secondly, the conditions to be tested are rated and these are known as 'item ratings' (IRs).

The listening sessions were conducted online by means of the software webMUSHRA [31]. Each participant completed two listening sessions with a duration of approximately 15–20 min each. These were separated by a break of at least one week so as to prevent over familiarisation of the stimuli which could lead to annoyance and bias in the ratings. Each listening session included an introduction page, a familiarisation page, a multiple stimuli BIR page and 20 single stimulus IR pages. Both of the two sessions were identical apart from the stimuli used in the single stimulus ratings.

On the introduction page participants were welcomed and asked to ensure that they were in a quiet space with headphones plugged into their device. The following instructions were given about the task to be completed:

> *In this experiment you will listen to various excerpts of music. For each excerpt you will be asked to rate your overall listening experience on a simple scale. In particular, you will be asked "How much do you enjoy listening to the following music item(s)" with possible answers ranging from "not at all" to "very much". When making your ratings you should take everything into account that you would normally in a real world scenario (e.g. your taste in music, the audio quality etc.).*

It should be noted that here the term 'overall listening experience' is expressed as a rating of 'enjoyment', as in previous implementations of the OLE method. This follows the assumption that for music consumption the overall experience can be represented by 'enjoyment' alone. Considering the definition of QoE [5] (based on [4]), which refers to the fulfilment of expectations and needs with respect to 'utility and/or enjoyment', this assumption is sensible as utility in this context is not relevant. For further discussions on the choice of question when assessing OLE please refer to Schoeffler [22] (pp. 35–42).

After the introduction, a familiarisation page allowed participants to play and rate four stimuli in order to adjust the volume of their device to a comfortable level and to practice using the interface. It was stated that once adjusted, the volume should not be changed during the remainder of the experiment. The four stimuli included one of each quality condition (as discussed in the following section) and were not used in the main rating pages. As with all of the rating pages, the order of the stimuli on the page was randomised. Ratings were made on a five-star Likert scale with labels of 'not at all', 'not a lot', 'neutral', 'quite' and 'very much', see Fig. 2. Before making a rating of an item, participants were required to listen to the item completely and before moving on to the next page, all items had to be rated. After the familiarisation page, participants made ratings of all of the stereo stimuli (ten items in total) presented on a single page. These ratings are the BIRs. Following this, single stimulus ratings were made for 20 stimuli which consisted of each content item at two quality levels. Moreover, each quality level appeared the same number of times in each session. Over the two sessions, participants therefore rated all stimuli (ten items by four conditions) by the single stimulus method. These ratings are the IRs. The allocation of stimuli to sessions was predetermined. To ensure that all combinations of quality levels for each content item were included, six configurations were needed and the assignment of these to participants was balanced.

Table 1 Overview of the content items used

Genre	Artist	Title	Duration (s)	Notes
Classical–choral	Bach	Komm, Jesu, Komm	21	Performed by The Sixteen for BBC Prom 42, 2016
Jazz–big band	Duke Ellington	Circle of Fourths	23	Performed by the National Youth Jazz Orchestra of Scotland for BBC Prom 28, 2016
Jazz–trumpet improv.	Duke Ellington	Lady Mac	24	Performed by the National Youth Jazz Orchestra of Scotland for BBC Prom 28, 2016
Folk	Hezekiah Jones	Borrowed Heart	25	Recorded for Weathervane Music's Shaking Through, Vol. 2, Ep. 4
Indie	Hop Along	Sister Cities	22	Recorded for Weathervane Music's Shaking Through, Vol. 4, Ep. 5
Electronic	La Big Vic	Musica	18	Recorded for Weathervane Music's Shaking Through Vol. 2, Ep. 3
Hip hop	Lushlife	Toynbee Suite	25	Recorded for Weathervane Music's Shaking Through Vol. 4, Ep. 8
Classical–orchestral	Prokofiev	Romeo and Juliet	23	Performed by the BBC National Orchestra of Wales for BBC Prom 16, 2016
Classical–orchestral	Schubert	Symphony No. 9	23	Performed by the BBC Philharmonic for BBC Prom 24, 2016
Pop	Steven A. Clarke	Bounty	20	Recorded for Weathervane Music's Shaking Through Vol. 4, Ep. 1
Folk*	Lea Thomas	Wild As You Are	20	Recorded for Weathervane Music's Shaking Through Vol. 8, Ep. 1
Classical–orchestral*	Schubert	Symphony No. 9	24	Performed by the BBC Philharmonic for BBC Prom 24, 2016
Classical–orchestral*	Tchaikovsky	Romeo and Juliet	16	Performed by the BBC Symphony Orchestra for BBC Prom 1, 2016
Indie*	The Tontons	Lush	23	Recorded for Weathervane Music's Shaking Through Vol. 5, Ep. 2

Starred items were used in the familiarisation stage only

Stimuli

Ten music items were used for the main rating sessions with an additional four being used for familiarisation pages, see Table 1. These spanned a range of genres and suitable phrases were selected that ranged in duration from 16–25 s (mean 21.9 s). The main selection criterion for these items was that they were available in formats that were suitable for the generation of binaural versions (due to the secondary objective of the experiment), i.e. captured with appropriate microphone techniques for the live classical and jazz performances and available as multitrack recordings for the popular items. Further criteria were that they were relatively broadband in nature, had a relatively wide stereo image and would elicit a range of preferences.

For each item four conditions were created: stereo, mono, 3.5 kHz low-pass filtered and binaural. All items were available as stereo mixes and these were used as the basis for the creation of the degraded mono and 3.5 kHz low-pass conditions. The mono items were created by passively downmixing the stereo items in accordance with ITU-R BS.775 [32]. The 3.5 kHz low-passed items were generated with a 5th-order Butterworth filter. A professional sound engineer experienced in mixing spatial audio assisted in the generation of the binaural items. More details on the production of these binaural items are presented in [28].

All stimuli had a 250 ms fade-in and fade-out applied and were presented as 44.1 kHz/16 bit WAV files. Additionally, a two-stage loudness alignment process was conducted to equalise the loudness of all stimuli. The first stage involved aligning all stereo items to a target loudness of − 18 LUFS in accordance with [33]. A target loudness of − 18 LUFS was chosen as such a level is more appropriate for mobile devices than the more typical − 23 LUFS [34]. Secondly, the remaining conditions for each item were aligned to the loudness of the stereo condition using the Glasberg and Moore loudness model applicable to time-varying sounds [35]. For each item, loudness values for each stereo channel were calculated individually and then these two values were averaged to produce a single loudness value. Furthermore, the model was applied without an outer ear transfer function stage as the stimuli were to be presented over headphones.

Participants

Participants were recruited through a variety of institutional mailing lists, social media, forums and participant recruitment websites, the aim being to recruit participants from a

range of backgrounds. In total, 58 participants completed all three sessions of the experiment. 45 participants with valid email addresses completed the online questionnaire but did not complete either of the listening sessions and seven participants completed the first listening session but did not complete the second listening session. This resulted in a total attrition rate of 47%. It should be mentioned that this is a higher attrition rate than one might find in laboratory experiments. High attrition rates in web-based studies have previously been reported and discussed elsewhere [36].

Results

Participant reliability

Suitability and reliability of participants was assessed by several means. Firstly, two participants self-reported that they did not have normal hearing and were therefore excluded from the analysis. In both listening sessions participants made basic item ratings of all ten stereo items. To assess participant reliability it was therefore possible to calculate the mean rating difference between the basic item ratings in each session. The mean BIR difference between the two sessions was 1.05, i.e. approximately one star on the rating scale, and the distribution around the mean was normal. Two participants were seen to have a BIR mean difference outside of 1.5 × the interquartile range, however, as these two outliers were close to the boundary of 1.5 × IQR (within 0.2 rating stars), it was decided that it was not necessary to exclude these participants from further analysis. Finally, the distribution of each participants' BIRs were checked in order to identify participants who may skew the results. Participants with a mode BIR at the extremes of the rating scale (i.e. those who chose 'not at all' or 'very much' most frequently) would potentially be limited in expressing improvements or deterioration due to the processing in comparison to their BIRs, i.e. floor and ceiling effects, as further discussed in [27]. It could therefore be expected that by including such participants, the difference in OLE ratings with respect to the different quality levels would be reduced and also that any correlations relating item ratings to quality levels would be weakened. Eight participants (14% of the 56 participants assessed) had a mode BIR of either 'not at all' or 'very much' and were therefore excluded from further analysis. The relatively high number of participants excluded at this stage is likely a result of the split in content between classical and jazz items and more contemporary items, coupled with the wide range of backgrounds of the participants. An alternative approach to excluding participants would be to individually select the items presented to each participant as in previous OLE experiments [27], however, this requires a larger pool of items to be rated than available for this study. To summarise, a total of ten participants (17%) were excluded and therefore data from 48 participants are used in the following analysis.

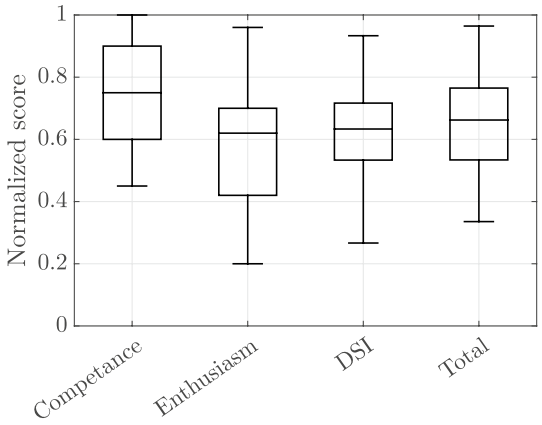


Fig. 3 Distribution of data from attitude related psychographic questions

Psychographic data

In this section, psychographic data from the remaining 48 participants are presented. For reference, the distribution of responses from all questions can be found in Appendix: Psychographic data.

In terms of demographics, the sample was predominantly male (69%), younger than 35 (61%) and educated to a university level (73%). The age range spanned from 18–25 (nine participants) to 66 or older (two participants).

With regards to work and hobbies, the sample was equally split between those who have work and hobbies related to audio technology and those who do not. The majority of participants listen to audio over headphones either everyday or often (77%). For binaural audio listening experience, half of the sample have some experience of listening to binaural audio, 25% have no experience and 25% are not sure. It could be assumed that those who are not sure are unfamiliar with the term 'binaural' and are therefore more likely to have no experience rather than some experience. If this is the case, the sample would be equally split between those who have no experience and those who have some experience of binaural audio listening. Finally, just less than half of the sample (46%) had not participated in listening tests previously.

Normalized scores for the competence, enthusiasm and DSI scales were calculated by assigning values of 1–5 to the Likert scale responses, summing these values (including inverting values for negative questions) and normalizing by the number of questions each scale contained. This resulted in values for each participant between 0 and 1 for the three measures. Additionally, a combined 'total' measure was

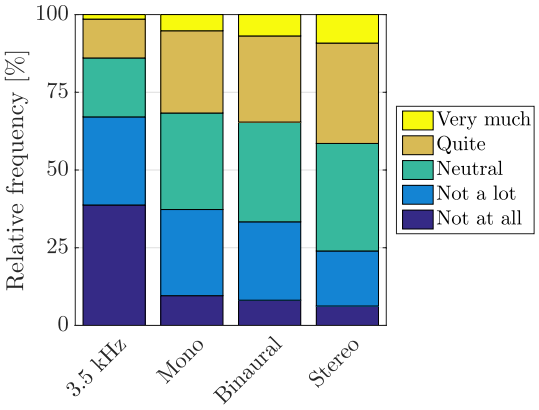


Fig. 4 Relative frequencies of item ratings grouped by processing

created by taking the mean of each participants' competence, enthusiasm and DSI values. Figure 3 shows the distribution of attitude values. It is seen that for all measures the median lies between 0.6 and 0.8, with competence having the highest median and enthusiasm the lowest. The smallest range in results is seen for competence (0.45 –1) and the largest for enthusiasm (0.2–0.96). Despite the skew to higher scores, the variation in attitudes is sufficient for the analysis presented in the following sections.

OLE analysis

As a full analysis of the OLE ratings with respect to the different processing conditions has previously been presented by Walton [28] in a paper with a narrower scope than this, the key points from the OLE analysis are summarised here.

The OLE ratings were made on a five-star Likert scale and as such could either be interpreted as ordinal data (from the labels) or interval data (from the number of stars). Typically it is recommended to use non-parametric statistics and median values for ordinal data, whereas with interval data it is possible to use parametric statistics and mean values. The choice of analysis for Likert-type data is well discussed in the literature and some prominent studies such as [37] advocate the use of either non-parametric or parametric analysis. Specific to the analysis of OLE, it was shown that there are only minor differences in effect sizes and statistical significance values when comparing non-parametric and parametric methods [38]. In the analysis of the OLE data presented here, the data are predominantly regarded as ordinal and as such non-parametric statistical techniques are used.

It is useful to gain an overview of the impact of the processing conditions on OLE, as presented in Fig. 4. When averaged over the different items it is seen that the 3.5 kHz condition has the lowest ratings followed by mono, binaural and stereo. Non-parametric Wilcoxon signed-rank tests are used to quantify the significance of the differences between these conditions and all comparisons reveal significant differences ($p < 0.05$). The timbral degradation introduced by a 3.5 kHz low-pass filter has much more of an impact on the ratings than either the mono or binaural processing. The small difference in ratings between mono, binaural and stereo suggest that, when averaged over participant and content, spatial processing has only a small affect on OLE. When comparing the stereo and binaural conditions, it is seen that binaural processing produces significantly lower ratings than stereo ($Z = -4.8$, $p < 0.001$), although the difference in ratings is small (an average of 0.2 stars).

Additionally, the ordering of processing conditions was investigated with respect to content group and listener group. The same order was found for both live and studio groups of content, as well as for multiple groups of participants; those whose OLE ratings were significantly influenced by spatial audio quality (five participants), those whose OLE ratings were significantly influenced by total audio quality (26 participants) and the sample as a whole.

Analysis of listener type

In this section, the influence of quality and content on OLE is determined for each participant. As suggested by Schoeffler and Herre [2], this is achieved by calculating Kendall rank correlation coefficients (Kendall's τ). Kendall's τ is a non-parametric statistic used to measure the ordinal association between two variables and results in a value ranging from -1 to $+1$. A value of -1 indicates perfect disagreement between the two variables, a value of 0 indicates that the two variables are independent and a value of $+1$ indicates perfect agreement between the two variables.

For each participant, four Kendall's τ values were calculated. To measure to what extent the content influences the OLE ratings, Kendall's τ was calculated from each participant's basic item ratings of all 10 stereo and basic item ratings ($\tau_{IR,BIR}$):

$$\tau_{IR,BIR} = cor_{\tau}(\mathbf{IR}, \mathbf{BIR}), \tag{1}$$

where **IR** and **BIR** are vectors of each participant's item ratings and basic items respectively. These vectors are sorted by item and processing and are therefore organised so that $\mathbf{IR}(i)$ and $\mathbf{BIR}(i)$ are ratings corresponding to the same item. To measure to what extent the timbral quality influences OLE ratings, Kendall's τ was calculated from each participant's item ratings associated with the 3.5 kHz and stereo conditions, and the associated timbral quality levels ($\tau_{IR,T}$):

$$\tau_{IR,T} = cor_{\tau}(\mathbf{IR}, \mathbf{T}), \tag{2}$$

where **T** is a vector containing the ranks of the timbral quality levels. The rank order of **T** is defined as: 3.5 kHz < stereo. $\mathbf{T}(i)$ therefore identifies the timbral quality level of $\mathbf{IR}(i)$ as either 3.5 kHz or stereo. To measure to what extent the

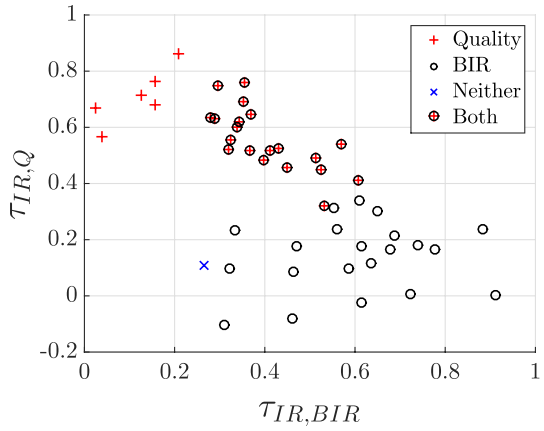


Fig. 5 Kendall's rank correlations between item rating and the two variables total quality level ($\tau_{IR,Q}$) and basic item rating ($\tau_{IR,BIR}$) for each participant. Each data point represents correlation values associated with one participant. Marker type indicates significant correlations ($p < 0.05$) between item ratings and the factors indicated in the legend, as determined by the Kendall's τ analysis. 'Quality' refers to overall quality level

spatial quality influences OLE ratings, Kendall's τ was calculated from each participant's item ratings associated with the mono and stereo conditions, and the associated spatial quality levels ($\tau_{IR,S}$):

$$\tau_{IR,S} = cor_{\tau}(\mathbf{IR}, \mathbf{S}), \tag{3}$$

where **S** is a vector containing the ranks of the spatial quality levels. The rank order of **S** is defined as: mono < stereo. **S**(i) therefore identifies the spatial quality level of **IR**(i) as either mono or stereo. It should be noted that the binaural condition is not included in the calculation of $\tau_{IR,S}$. One requirement for Kendall's τ analysis is that there is a monotonic relationship between the two variables. As such, it was decided to exclude the binaural quality level from the analysis as this quality level was not consistently rated between the mono and stereo quality levels when considering the results on a participant by participant basis. In other words, participants' ratings would not necessarily reflect the rank order of mono < binaural < stereo, thus breaking the assumption of a monotonic relationship between **IR** and **S**. Finally, to measure to what extent the overall quality influences OLE ratings, Kendall's τ was calculated from each participant's item ratings associated with the 3.5 kHz, mono and stereo conditions, and the associated quality levels ($\tau_{IR,Q}$):

$$\tau_{IR,Q} = cor_{\tau}(\mathbf{IR}, \mathbf{Q}), \tag{4}$$

where **Q** is a vector containing the ranks of the overall quality levels. The rank order of **Q** is defined as: 3.5 kHz < mono < stereo. **S**(i) therefore identifies the overall quality level of **IR**(i) as either 3.5 kHz, mono or stereo.

Fig. 5 presents a scatter plot of $\tau_{IR,Q}$ values versus $\tau_{IR,BIR}$ values for each participant. In this plot (and the subsequent correlation plots), each data point represents correlation values associated with one participant. Furthermore, the marker type indicates whether each participant's correlation values are significant ($p < 0.05$) for the correlations in question, as calculated from the Kendall's τ analysis. For example, in Fig. 5, each participant's item ratings can be significantly correlated with the overall quality level (red plus), their basic item ratings (black circle), neither overall quality level nor BIRs (blue ×), or both overall quality level and BIRs (black circle filled with red plus), as determined by the Kendall's τ calculations of $\tau_{IR,Q}$ and $\tau_{IR,BIR}$. Those participants with a high $\tau_{IR,Q}$ value are heavily influenced by the technical audio quality when making OLE ratings and those participants with a high $\tau_{IR,BIR}$ value are heavily influenced by the content when making OLE ratings. Applying Pearson's correlation to the data reveals a strong negative correlation between the pairs of τ values ($r = -0.63$).[2] In other words, participants who are more influenced by technical audio quality are less influenced by content and *vice versa*. This is in line with results presented by Schoeffler and Herre [2], who reported that a continuum exists that describes to what extent a listener's OLE ratings are influenced by technical audio quality and content. From Fig. 5 it is also apparent that some listeners are weakly influenced by both quality and content, represented by the data points at low values on both axes.

To estimate in what way listeners affected by timbral and spatial quality individually were also affected by overall quality, scatter plots of $\tau_{IR,T}$ versus $\tau_{IR,Q}$ and $\tau_{IR,S}$ versus $\tau_{IR,Q}$ are examined, Fig. 6. Pearson's correlation reveals a very strong positive correlation ($r = 0.99$) between $\tau_{IR,T}$ and $\tau_{IR,Q}$ and a strong positive correlation ($r = 0.72$) between $\tau_{IR,S}$ and $\tau_{IR,Q}$. This shows that there is a greater correlation between timbral quality and overall quality compared to spatial quality and overall quality, and this is expected given the OLE ratings presented in Fig. 4. The stronger influence of timbral quality compared to spatial quality is also coherent with previous studies such as Rumseyet al. [40], however, a direct comparison cannot be made as total bandwidth between the timbral and spatial degradations were not matched in this study.

To assess if participants who are influenced by timbral quality are also influenced by spatial quality, $\tau_{IR,T}$ versus $\tau_{IR,S}$ is plotted, Fig. 7. Pearson's correlation reveals a strong positive correlation ($r = 0.64$) which indeed suggests that participants who are influenced by timbal quality are more likely to be influenced by spatial quality. It is apparent from Fig. 7 that only a small number of participants are significantly

[2] For effect sizes in behavioural research, Cohen's conventions are typically used [39]. These state that an r of |0.1| represents a 'small' effect size, |0.3| represents a 'medium' effect size and |0.5| represents a 'large' effect size.

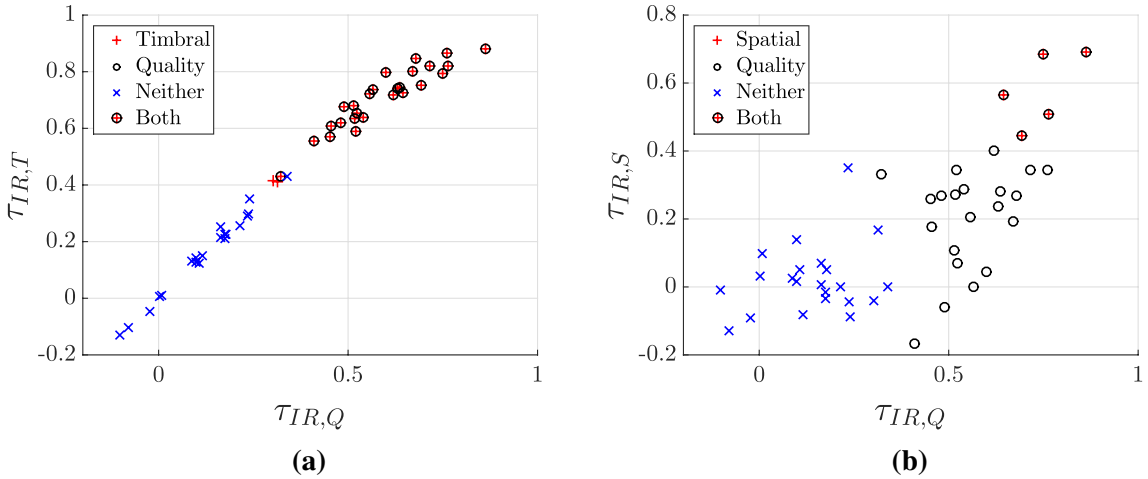


Fig. 6 Kendall's rank correlations between item rating and the two variables timbral quality level ($\tau_{IR,T}$) and total quality level ($\tau_{IR,Q}$) (**a**) and the two variables spatial quality level ($\tau_{IR,S}$) and total quality level ($\tau_{IR,Q}$) (**b**). Each data point represents correlation values associated with one participant. Marker type indicates significant correlations ($p < 0.05$) between item ratings and the factors indicated in the legend, as determined by the Kendall's τ analysis. 'Timbral', 'Quality' and 'Spatial' refer to the timbral, overall and spatial quality levels respectively

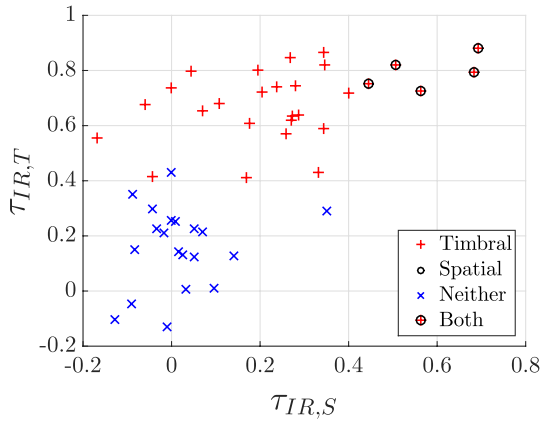


Fig. 7 Kendall's rank correlations between item rating and the two variables timbral quality level ($\tau_{IR,T}$) and spatial quality level ($\tau_{IR,S}$) for each participant. Each data point represents correlation values associated with one participant. Marker type indicates significant correlations ($p < 0.05$) between item ratings and the factors indicated in the legend, as determined by the Kendall's τ analysis. 'Timbral' and 'Spatial' refer to timbral and spatial quality levels respectively

influenced by spatial quality (five) and all of these are also significantly influenced by timbral quality. Furthermore, there are some participants who are significantly influenced by timbral quality but have very low correlations with spatial quality. It could therefore be said that participants who are significantly influenced by spatial quality will typically be influenced by timbral quality, but this is not the case in reverse.

Interaction between psychographic variables and listener type

The interaction between the psychographic variables and the measures of listener type is now investigated. Except for gender, all of the psychographic variables are measured on an ordinal or continuous scale and as such Kendall's τ can be used to investigate correlations between the psychographic variables and measures of listener type (τ_Q, τ_T, τ_S and τ_{BIR}), see Table 2. As gender is a dichotomous variable, a point-biserial correlation is used instead. Spearman's rank correlation was also used to verify the significant correlations. It is seen that the variables age, education and headphone usage do not show any significant correlations with the measures of listener type. The remaining variables on the other hand, all show significant correlations with one measure of listener type or more. The variable competence shows the strongest correlation with the measures $\tau_{IR,Q}$ and $\tau_{IR,T}$. For $\tau_{IR,S}$ and $\tau_{IR,BIR}$, the strongest correlation is with gender. Care should be taken when interpreting this result however, and indeed the other gender correlations, as the sample was not equally stratified by gender. For example, with regards to work and hobbies no females answered 'I study or work mainly in the field of audio technology' compared to 12 males. It therefore cannot be assumed that it is gender itself that leads to the significant correlations seen. Excluding gender, the strongest correlation with the measures $\tau_{IR,S}$ and $\tau_{IR,BIR}$ is also competence. As well as the variable competence, variables total attitude (which includes competence), work/hobbies and enthusiasm all show correlations above $\tau = 0.4$ for $\tau_{IR,Q}$.

To predict measures of listener type from the psychographic variables, stepwise multiple regressions were performed for

Table 2 Kendall's τ correlation and significance between psychographic variables and measures of listener type $\tau_{IR,Q}$, $\tau_{IR,T}$, $\tau_{IR,S}$ and $\tau_{IR,BIR}$

	$\tau_{IR,Q}$	$\tau_{IR,T}$	$\tau_{IR,S}$	$\tau_{IR,BIR}$
Gender*	**$r = 0.487$, $p < 0.001$**	**$r = 0.484$, $p < 0.001$**	**$r = 0.335$, $p = 0.020$**	**$r = -0.416$, $p = 0.003$**
Age	$\tau = 0.078$, $p = 0.478$	$\tau = 0.080$, $p = 0.466$	$\tau = 0.128$, $p = 0.247$	$\tau = 0.006$, $p = 0.955$
Education	$\tau = 0.063$, $p = 0.581$	$\tau = 0.091$, $p = 0.424$	$\tau = -0.061$, $p = 0.594$	$\tau = 0.059$, $p = 0.607$
Work/hobbies	**$\tau = 0.442$, $p < 0.001$**	**$\tau = 0.454$, $p < 0.001$**	**$\tau = 0.270$, $p = 0.018$**	**$\tau = -0.352$, $p = 0.002$**
Headphones usage	$\tau = 0.207$, $p = 0.068$	$\tau = 0.199$, $p = 0.080$	$\tau = 0.165$, $p = 0.146$	$\tau = -0.130$, $p = 0.251$
Binaural exp.	**$\tau = 0.243$, $p = 0.028$**	**$\tau = 0.256$, $p = 0.021$**	$\tau = 0.188$, $p = 0.092$	$\tau = -0.181$, $p = 0.103$
Prev. listening tests	**$\tau = 0.225$, $p = 0.049$**	**$\tau = 0.262$, $p = 0.022$**	$\tau = 0.079$, $p = 0.487$	**$\tau = -0.265$, $p = 0.020$**
Competence	**$\tau = 0.537$, $p < 0.001$**	**$\tau = 0.549$, $p < 0.001$**	**$\tau = 0.302$, $p = 0.003$**	**$\tau = -0.397$, $p < 0.001$**
Enthusiasm	**$\tau = 0.440$, $p < 0.001$**	**$\tau = 0.468$, $p < 0.001$**	**$\tau = 0.258$, $p = 0.012$**	**$\tau = -0.324$, $p = 0.002$**
DSI	**$\tau = 0.322$, $p = 0.002$**	**$\tau = 0.353$, $p = 0.001$**	$\tau = 0.145$, $p = 0.156$	**$\tau = -0.208$, $p = 0.042$**
Total attitude	**$\tau = 0.481$, $p < 0.001$**	**$\tau = 0.523$, $p < 0.001$**	**$\tau = 0.271$, $p = 0.007$**	**$\tau = -0.346$, $p = 0.001$**

*For gender, a point-biseiral correlation was used. Significant correlations are highlighted in bold

each measure. The independent variables in the regressions were the significant psychographic variables associated with each measure (presented in Table 2), excluding gender and also total attitude (due to possible multicollinearity problems with the variables that make up total attitude). The relevant assumptions related to multiple regression analysis were checked including independence of residuals, linear relationships between the dependent and independent variables, homoscedasticity, multicollinearity issues and normal distribution of residuals.

For all four measures of listener type ($\tau_{IR,Q}$, $\tau_{IR,T}$, $\tau_{IR,S}$ and $\tau_{IR,BIR}$), competence was the only variable that added significantly to the prediction and, as such, all other variables were excluded from the model. The specific significance values and model coefficients are listed below.

$$\tau_{IR,Q} = -0.418 + (1.075 \times \text{competence})$$
$$F(1, 46) = 42.506, p < 0.0005, R = 0.693, R^2 = 0.480$$

$$\tau_{IR,T} = -0.459 + (1.242 \times \text{competence})$$
$$F(1, 46) = 43.652, p < 0.0005, R = 0.698, R^2 = 0.487$$

$$\tau_{IR,S} = -0.252 + (0.550 \times \text{competence})$$
$$F(1, 46) = 10.991, p = 0.002, R = 0.439, R^2 = 0.193$$

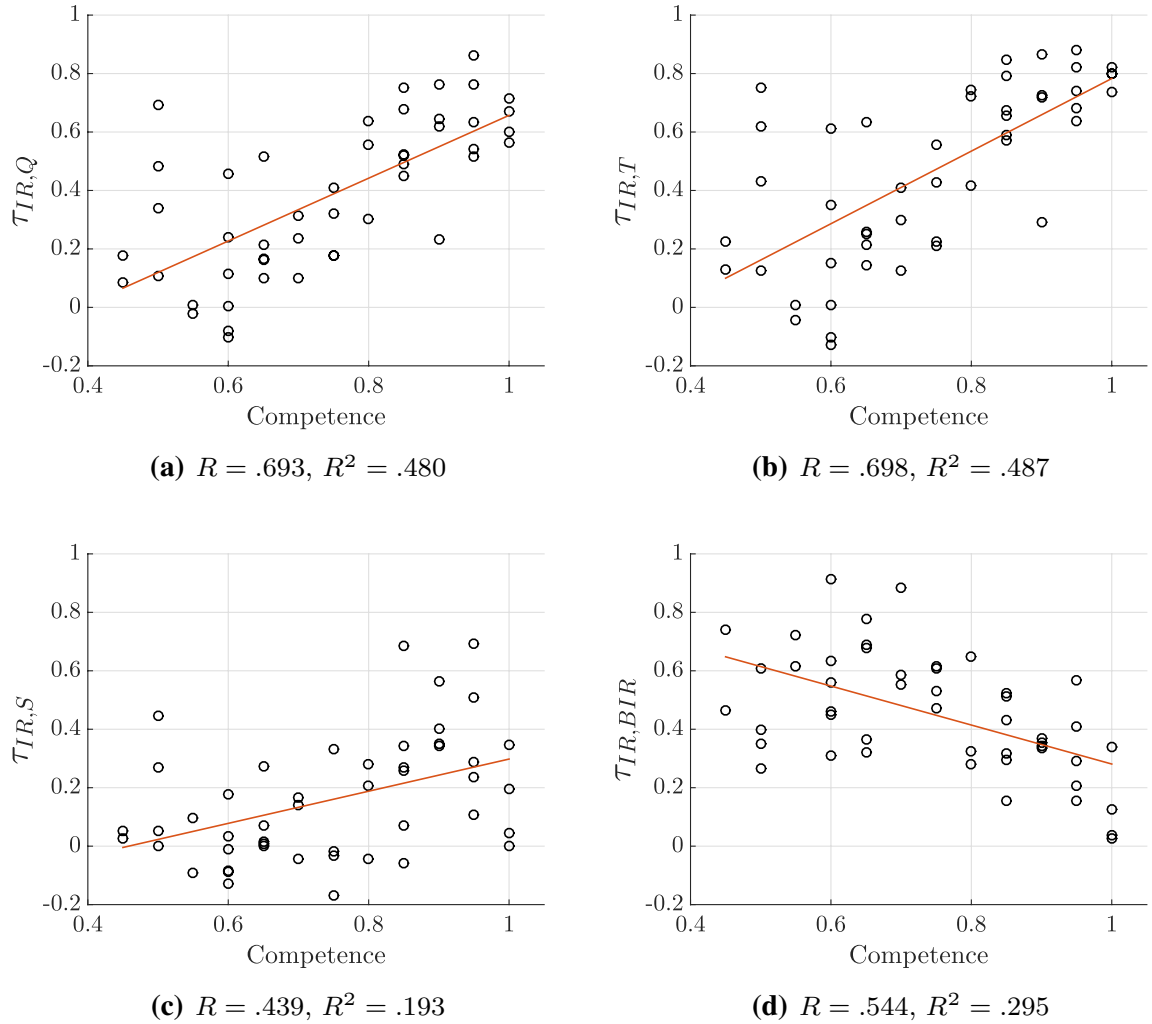
$$\tau_{IR,BIR} = 0.948 - (0.667 \times \text{competence})$$
$$F(1, 46) = 19.293, p < 0.0005, R = 0.544, R^2 = 0.295$$

From the above values, it is seen that competence explains 48% of the variance in $\tau_{IR,Q}$, 49% of the variance in $\tau_{IR,T}$, 19% of the variance in $\tau_{IR,S}$ and 30% of the variance in $\tau_{IR,BIR}$. The effect sizes (R values) can all be classified as strong, with the exception of the correlation between $\tau_{IR,S}$ and competence which can be classified as moderate. The correlations between the measures of listener type and competence are presented in graphical form in Fig. 8. When looking at the plot of $\tau_{IR,Q}$ versus competence, it is seen that participants with a high competence score (> 0.8) have relatively similar $\tau_{IR,Q}$ values (within 0.4 of each other), with the exception of one participant. On the other hand, participants with a low competence score (≤ 0.6) show a larger range in $\tau_{IR,Q}$ values. That is to say, participants who have high competence scores are typically highly influenced by technical audio quality when making OLE ratings, however the opposite is less certain to be true for participants with low competence scores. On the other hand, when looking at the plot of $\tau_{IR,S}$ versus competence it is seen that the largest range of $\tau_{IR,S}$ values are for participants with high competence values. In other words, it is hardest to predict how much a listener will be influenced by spatial audio quality for participants who have high competence scores. Finally, when looking at the plot of $\tau_{IR,BIR}$ versus competence, a negative correlation is seen which shows that participants with high competence scores are typically less influenced by the content than participants with low competence scores, and this is expected given the previous results.

Complementary analysis

The analyses presented thus far have been predominantly based on correlation values, as calculated to identify listener type. To support the conclusions drawn in the above sections it is beneficial to also provide analysis based on the raw item ratings. Specifically, in this section we aim to support the above conclusion that the attitudinal measure 'competence' is a significant predictor of listener type by conducting an analysis of variance on the raw item ratings.

A three-way mixed ANOVA was conducted on normalized IR data with overall quality level (four levels) and content (ten levels) as within-subject factors, and competence (two levels) as a between-subject factor. Note that this is a parametric analysis and therefore the OLE ratings are considered as interval data, as discussed previously. The IR data was normalized to the BIR data by subtracting participants' BIRs from their IRs, where the BIRs are from the



(a) $R = .693,\ R^2 = .480$ **(b)** $R = .698,\ R^2 = .487$

(c) $R = .439,\ R^2 = .193$ **(d)** $R = .544,\ R^2 = .295$

Fig. 8 Correlations between measures of listener type and the psychographic variable competence with regression lines plotted. Each data point represents correlation values associated with one participant

corresponding session. The normalized IR data therefore takes into account the degree of liking of the content and is a measure of the deviation between participants' basic item ratings and item ratings. To prepare the competence data for the ANOVA, it was necessary to transform it from continuous data to categorical data. This was achieved by dichotomising the competence scores into a low competence group (22 participants) and a high competence group (26 participants), split around the mean.

Prior to analysis the assumptions underlying the mixed ANOVA were checked, namely, normality for each combination of the within-subject and between-subject factors, homogeneity of variances for each combination of the groups and sphericity. The factor combinations were largely normally distributed with predominantly homogenous variances across between-subject factors. However, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the within-subject factor of overall quality level ($\chi^2(5) = 29.1, p < 0.001$) and therefore a Greenhouse-Geisser correction was applied.

By studying the interaction between overall quality level and competence, it is possible to assess if the impact of overall quality on normalized item ratings (i.e. listener type) is influenced by competence. The interaction between overall quality level and competence was found to have a significant influence on the normalized item ratings [$F(2.20, 101) = 20.4, p < 0.001$]. Furthermore, a partial eta-squared value of $\eta_p^2 = 0.307$ indicates a large effect. Figure 9 represents this interaction graphically, by plotting normalized item ratings (averaged over content) with respect to both overall quality level and competence group. It is seen that participants in the low competence group are only mildly influenced by the overall quality level. On the other hand, participants in the high competence group are significantly

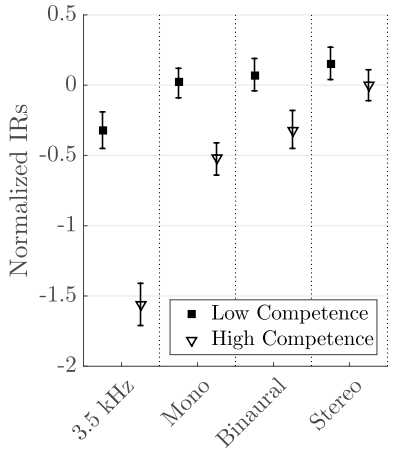


Fig. 9 Normalized item ratings (averaged over content) with respect to both overall quality level and competence group. Error bars show 95% confidence intervals

more influenced by the overall quality level, thus supporting the results from the previous sections.

Discussion

The aim of this study was to investigate correlations between psychographic variables and the influence of technical audio quality on overall listening experience. The first stage of this was to evaluate to what extent each participant was influenced by technical audio quality and the content when making OLE ratings. As with previous studies [2], a negative correlation was found between the influence of quality and the influence of content on OLE ratings. Participants who are more influenced by technical audio quality are generally less influenced by content and *vice versa*. In previous studies labels of 'audio quality likers' and 'song likers' were used to describe this range in participants [1]. When looking at the influences of timbral audio quality and spatial audio quality on OLE, only five out of 48 participants were significantly influenced by spatial audio quality compared to 28 who were significantly influenced by timbral audio quality. All of those who were significantly influenced by spatial audio quality were significantly influenced by timbral audio quality. A strong positive correlation between the influence of timbral audio quality and the influence of spatial audio quality on OLE ratings was seen, which suggests that participants who are influenced by one aspect of quality are likely to be influenced by other aspects of quality.

Interactions between psychographic variables and listener type were studied by means of correlation and regression analysis. The psychographic variables that showed significant correlations with the influence of technical audio quality on OLE ratings included work and hobbies, binaural experience, previous listening tests, competence, enthusiasm, innovativeness and total attitude towards audio technology. To predict the influence of technical audio quality on OLE a multiple regression analysis was performed. The only psychographic variable that added significantly to the prediction was the attitudinal measure competence and indeed a strong correlation between competence and the influence of technical audio quality on OLE was seen ($R = 0.693$). This was also supported by an analysis of variance on the raw OLE ratings. This result suggests that, out of the psychographic variables studied in this experiment, the measure competence is the most useful for predicting to what extent a listener will be influenced by degradations in technical audio quality. This measure consists of four questions and is thus a practical way to quickly assess how a participant may respond to different levels of technical audio quality. Applications of this knowledge could include adapting the technical audio quality requirements of a product or service to the potential users in a more educated way, improving quality prediction models and also using the competence questionnaire presented here for participant recruitment purposes in subjective evaluations. Typically in subjective evaluations of audio, data collected about participants includes professionalism and previous number of listening tests, however this study shows that gathering data about the attitudinal measure competence could be more worthwhile in some cases.

Additional human influence factors

Despite there being a strong correlation between competence and influence of technical audio quality on OLE, a large variance around the regression line was seen. As this experiment was limited in the number of human influence factors studied, further studies should look for additional factors that help explain some of the variance not described by the variables used in this experiment. The basic approach used in this study, in which each participant responded to a structured psychographic questionnaire before rating the stimuli in a listening session, has scope for significant extension to encompass other human influence factors. This form of pre-listening approach could be applied to any human influence factor that is reliably signalled through questionnaire responses, including prevailing effects of emotion and mood, attention level and participative motivation. However, such states are generally subject to potentially significant transient variation from a prevailing level and the reliability of the pre-listening approach may be compromised when investigating the effect of such factors. Alternative measures, perhaps using secondary techniques of observation through sensors or behavioural coding, might be useful as an alternative or supplemental source of participant state data.

It is also likely that the relative influence of content and technical audio quality on OLE is context specific, in

addition to being user specific. It would therefore be worthwhile investigating how these groups of influence factors interrelate with regards to OLE. For instance, it may be the case that some participants are heavily influenced by technical audio quality in a home context, but not in a mobile listening context. A straightforward approach to scaling up the study method used in this paper, would be to implement the psychographic and rating phases for a mobile device, and conduct the study with participants in a diverse range of contexts and environments. This approach to investigating such relationships harnesses the benefits of the relative speed and simplicity of the method and, therefore, its capacity for straightforward replication.

Conclusion

In this article, the influence of human factors on overall listening experience was studied by means of a web-based experiment. Previous work had shown that the relative influence of content and technical quality on overall listening experience is very listener dependent. In this study, we expanded on this previous research by investigating correlations between such listener types and a range of psychographic variables. It was shown that listener type is significantly correlated with multiple psychographic variables and that the attitudinal measure 'competence' is the most suitable variable to be used as a predictor of listener type.

As well as having direct applications such as tailoring systems and services to the needs of users, the results presented here add to the understanding of how human factors can influence quality of experience. Human factors are perhaps the least studied group of influence factors, however, their understanding is critical to providing a high quality of experience for the user.

Acknowledgements The authors would like to thank the reviewers for their detailed comments and suggestions for the manuscript. Additionally, the authors would like to thank Tom Parnell for mixing the audio content, Linda Beilig for translating the technical affinity questionnaire and Michael Schoeffler for discussions on the method. This work was supported by an Engineering and Physical Sciences Research Council (EPSRC) industrial CASE studentship [grant number EP/L505560/1] in partnership with the British Broadcasting Corporation.

Compliance with ethical standards

Conflict of interest statement On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix: psychographic data

In this appendix, the distribution of responses from all questions in the psychographic survey are presented (Figs. 10, 11, 12, 13, 14, 15, 16).

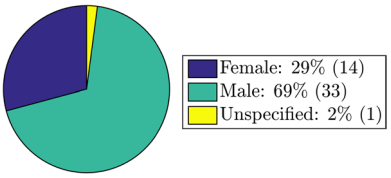


Fig. 10 Distribution of gender

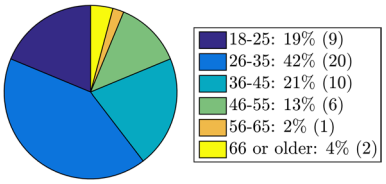


Fig. 11 Distribution of age

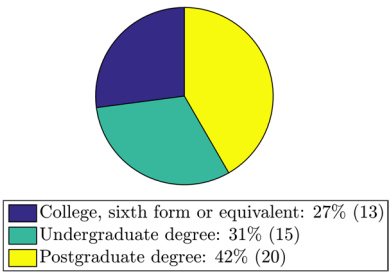


Fig. 12 Distribution of education level

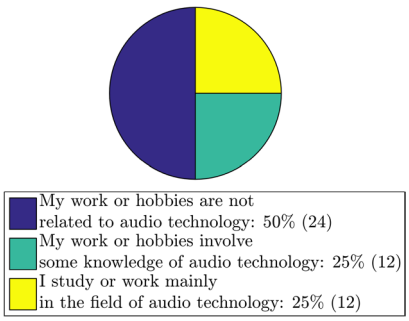


Fig. 13 Distribution of work and hobbies

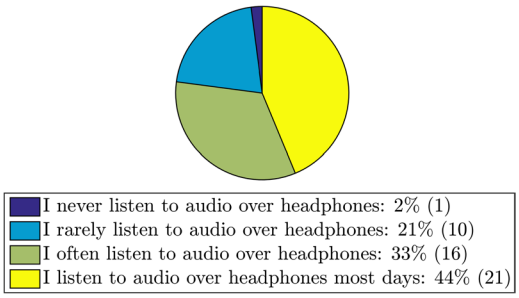


Fig. 14 Distribution of headphone usage

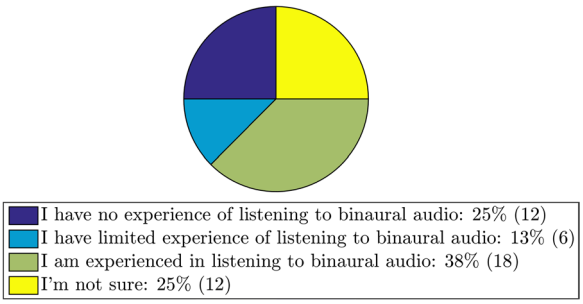


Fig. 15 Distribution of binaural experience

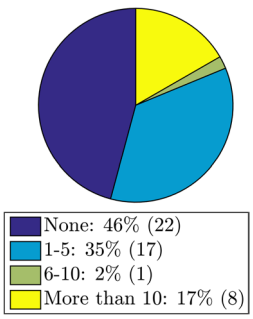


Fig. 16 Distribution of previous listening tests

References

1. Schoeffler M, Herre J (2013) About the impact of audio quality on overall listening experience. In: Proceedings of sound and music computing Conference, Stockholm, pp 48–53. Available at http://smcnetwork.org/node/1752. Accessed 03 Nov 2017
2. Schoeffler M, Herre J (2014) About the different types of listeners for rating the overall listening experience. In: Proceedings of sound and music computing Conference 2014, Athens. Available at http://speech.di.uoa.gr/ICMC-SMC-2014/index.php. Accessed 03 Nov 2017
3. OED Online (2017) Psychographics. Oxford University Press. https://en.oxforddictionaries.com/definition/psychographics. Accessed 03 Nov 2017
4. Le Callet P, Möller S, Perkis A (eds) (2013) Qualinet white paper on definitions of quality of experience, Version 1.2. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne
5. Möller S, Raake A (eds) (2014) Quality of experience: advanced concepts, applications and methods. Springer, Berlin
6. Jumisko-Pyykkö S (2011) User-centered quality of experience and its evaluation methods for mobile television. PhD thesis, Tampere University of Technology
7. Jumisko-Pyykkö S, Häkkinen J (2008) Profiles of the evaluators: impact of psychographic variables on the consumer-oriented quality assessment of mobile television. In: Proceedings of IS&T/SPIE's International Symposium on Electronic imaging: science and technology: multimedia on mobile devices
8. Wechsung I, Schulz M, Engelbrecht KP, Niemann J, Möller S (2011) All users are (not) equal—the influence of user characteristics on perceived quality, modality choice and performance. In: Proceedings of the Paralinguistic information and its integration in spoken dialogue systems workshop. Springer, New York, pp 175–186
9. Rainer B, Waltl M, Cheng E, Shujau M, Timmerer C, Davis S, Burnett I, Ritz C, Hellwagner H (2012) Investigating the impact of sensory effects on the quality of experience and emotional response in web videos. In: 2012 Fourth International Workshop on Quality of multimedia experience, pp 278–283
10. Arndt S, Antons JN, Schleicher R, Möller S, Curio G (2012) Perception of low-quality videos analyzed by means of electroencephalography. In: 2012 Fourth International Workshop on Quality of multimedia experience, pp 284–289
11. Reiter U, Moor KD (2012) Content categorization based on implicit and explicit user feedback: combining self-reports with eeg emotional state analysis. In: 2012 Fourth International Workshop on Quality of multimedia experience, pp 266–271
12. Ryan RM, Deci EL (2000) Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 25(1):54–67
13. Sackl A, Masuch K, Egger S, Schatz R (2012) Wireless vs. wireline shootout: how user expectations influence quality of experience. In: 2012 Fourth International Workshop on Quality of multimedia experience, pp 148–149
14. Staelens N, Van den Broeck W, Pitrey Y, Vermeulen B, Demeester P (2012) Lessons learned during real-life qoe assessment. In: 10th European Conference on Interactive TV and video, Proceedings, Ghent University, Department of Information technology, pp 1–4
15. Sackl A, Schatz R, Raake A (2017) More than I ever wanted or just good enough? User expectations and subjective quality perception in the context of networked multimedia services. Qual User Exp 2(1):3
16. Rumsey F, Zielinski S, Kassier R, Bech S (2005) Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences. J Acoust Soc Am 117(6):3832–3840
17. Quintero RM, Raake A (2012) Is taking into account the subjects degree of knowledge and expertise enough when rating quality? Fourth International Workshop on Quality of multimedia experience, QoMEX 2012. Melbourne, pp 194–199
18. Kim S, Bakker R, Ikeda M (2016) Timbre preferences of four listener groups and the influence of their cultural backgrounds. In: Audio Engineering Society Convention 140
19. Olive S, Welti T, McMullin E (2014) The influence of listeners' experience, age, and culture on headphone sound quality preferences. In: Audio Engineering Society Convention 137
20. Ebem DU, Beerends JG, Van Vugt J, Schmidmer C, Kooij RE, Uguru JO (2011) The impact of tone language and non-native language listening on measuring speech quality. J Audio Eng Soc 59(9):647–655
21. Schinkel-Bielefeld N, Jiandong Z, Yili Q, Leschanowsky AK, Shanshan F (2017) Is it harder to perceive coding artifact in

foreign language items?—A study with mandarin chinese and german speaking listeners. In: Audio Engineering Society Convention 142
22. Schoeffler M (2017) Overall listening experience—A new approach to subjective evaluation of audio. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg
23. Schoeffler M, Edler B, Herre J (2013) How much does audio quality influence ratings of overall listening experience? In: Proceedings of the 10th Annual Symposium on Computer Music Multidisciplinary Research
24. Schoeffler M, Adami A, Herre J (2014a) The influence of up- and down-mixes on the overall listening experience. In: Audio Engineering Society Convention 137
25. Schoeffler M, Conrad S, Herre J (2014b) The influence of the single/multi-channel-system on the overall listening experience. In: Proceeding of the AES 55th Conference on spatial audio, Helsinki
26. Schoeffler M, Silzle A, Herre J (2017) Evaluation of spatial/3D audio: basic audio quality vs. quality of experience. IEEE J-STSP 11(1):75–88
27. Schoeffler M, Herre J (2016) The relationship between basic audio quality and overall listening experience. J Acoust Soc Am 140(3):2101–2112
28. Walton T (2017) The overall listening experience of binaural audio. In: Proceeding of the 4th International Conference on spatial audio (ICSA 2017), Graz
29. Goldsmith RE, Hofacker CF (1991) Measuring consumer innovativeness. J Acad Market Sci 19(3):209–221
30. Karrer K, Glaser C, Clemens C, Bruder C (2009) Technikaffinität erfassen der fragebogen TA-EG [Assessing technical affinity - the questionnaire TA-EG]. In: Lichtenstein A, Stößel C, Clemens C (eds) Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme. VDI Verlag GmbH, Düsseldorf, pp 196–201
31. Schoeffler M, Stöter FR, Edler B, Herre J (2015) Towards the next generation of web-based experiments: A case study assessing basic audio quality following the ITU-R recommendation BS.1534 (MUSHRA). In: 1st Web Audio Conference, Paris
32. ITU-R (2012) BS.775-3 multichannel stereophonic sound system with and without accompanying picture. International Telecommunication Union
33. ITU-R (2015) BS.1770-4 algorithms to measure audio programme loudness and true-peak audio level. International Telecommunication Union
34. AES (2015) AES TD1004.1.15-10 Recommendation for loudness of audio streaming and network file playback. Audio Engineering Society Technical Document
35. Glasberg BR, Moore BCJ (2002) A model of loudness applicable to time-varying sounds. J Audio Eng Soc 50(5):331–342
36. Mason W, Suri S (2012) Conducting behavioral research on amazon's mechanical turk. Behav Res Methods 44(1):1–23
37. Norman G (2010) Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ 15(5):625–632
38. Schoeffler M, Gernert JL, Neumayer M, Westphal S, Herre J (2015) On the validity of virtual reality-based auditory experiments: a case study about ratings of the overall listening experience. Virtual Real 19(3–4):1–20
39. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd Edition). Routledge, London
40. Rumsey F, Zielinski S, Kassier R, Bech S (2005) On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. J Acoust Soc Am 118(2):968–976